Three steps ahead

Yuval Heller

University of Oxford, Department of Economics

13. June 2012

Three Steps Ahead (August 17, 2012)

*Yuval Heller**

Address: Nuffield College and Dept. of Economics, New Road, Oxford, OX1 1NF, United Kingdom. Email: yuval26@gmail.com or yuval.heller@economics.ox.ac.uk.

Abstract

Experimental evidence suggest that people only use a few iterations of strategic reasoning, and that some people systematically use less iterations than others. In this paper, we present a novel evolutionary foundation for these stylized facts. In our model, agents interact in a finitely repeated Prisoner's Dilemma, and each agent is characterized by the number of steps he thinks ahead. When two agents interact, each of them has an independent probability to observe the opponent's type. We show that if this probability is not too close to 0 or 1, then the evolutionary process admits a unique stable outcome, in which the population includes a mixture of "naive" agents who think 1 step ahead, and "moderately-sophisticated" agents who think up to 3 steps ahead.

Keywords: Indirect evolution, evolutionary stability, cognitive hierarchy, bounded forward-looking, Prisoner's Dilemma, Cooperation. JEL Classification: C73, D03.

1 Introduction

Experimental evidence suggest that in new strategic interactions most people only use a few iterations of strategic reasoning. This stylized fact is observed in different forms in various contexts. First, when playing long finite games, people only look a few stages ahead and use backward induction reasoning to a limited extent. For example, players usually defect only at the last couple of stages when playing finitely-repeated Prisoner's Dilemma, (see, e.g., Selten and Stoecker (1986)) and "Centipede" games (McKelvey and Palfrey (1995); Nagel and

*I would also like to express my deep gratitude to Itai Arieli, Vince Crawford, Ariel Rubinstein, Peyton Young, and seminar participants at University of Birmingham, University of Oxford and University College London, for many useful comments, discussions and ideas.

Tang (1998)), and when interacting in sequential bargaining, players ignore future bargaining opportunities that are more than 1-2 steps ahead (Neelin, Sonnenschein, and Spiegel (1988); Johnson, Camerer, Sen, and Rymon (2002)). Second, when facing iteratively dominated strategies, almost everyone make the first iteration (not playing a dominated action), many do the second iteration - assume that their opponent does not play dominated strategies, a few make the third iteration, and further iterations are rare (see, e.g., Costa-Gomes, Crawford, and Broseta (2001); Rapoport and Amaldoss (2004); Costa-Gomes and Crawford (2006)). Third, according to the models of cognitive hierarchy (or level-k), most players best respond to a belief that others use at most two iterations of strategic reasoning (see, e.g., Stahl and Wilson (1994); Nagel (1995); Ho, Camerer, and Weigelt (1998); Bosch-Domenech, Montalvo, Nagel, and Satorra (2002); Camerer, Ho, and Chong (2004); Crawford and Iriberri (2007)).

A second stylized fact is the heterogeneity of the population: some people systemically use less iterations than others (Chong, Camerer, and Ho (2005); Costa-Gomes and Crawford (2006); Hyndman, Terracol, and Vaksmann (2012)). These observations raise two related evolutionary puzzles. The first puzzle is why people only use few steps. Experimental evidence suggest that using more iterations is only unintuitive but not computationally complex (at-least in simple games): with appropriate guidance and feedback players can learn to use many iterations in a given game (Crawford (2008); Camerer (2003, Section 5.3.5)). In many games, being able to do one more step than the opponent gives a substantial advantage. As the cognitive cost of an additional level is moderate, it is puzzling why there was not an "arms race" in which people learn to use more strategic iterations throughout the evolutionary process ("red queen effect", see Robson (2003)).

The second puzzle is how the "naive" people, who systematically use less iterations than the more "sophisticated" agents, survived the evolutionary process. At first glance, it seems that sophisticated agents would outperform naive agents due to the benefit of thinking one level ahead. *In this paper we present an evolutionary model that explains both puzzles and yields a unique sharp prediction: an heterogeneous population of naive agents and moderately-sophisticated agents, in which everyone uses only 1-3 strategic iterations.* Our model focuses on bounded forward-looking in repeated Prisoner's Dilemma. We believe that it can also shed light on other forms of bounded iterative reasoning.

Following the "indirect evolutionary approach" (Güth and Yaari (1992)) we present a reduced-form static analysis for a dynamic process that describes the evolution of types in a large population of agents.[1] This process can be interpreted in two different ways: (1)

[1] The indirect approach was mainly used to study evolution of preferences, and it is related the literature on strategic delegation (e.g., Fershtman, Judd, and Kalai (1991)). Following, Stahl (1993); Stennek (2000); Frenkel, Heller, and Teper (2012), we apply it to analyze evolution of cognitive biases.

	C	D
C	A,A	A+1,0
D	A+1,0	1,1

Tab. 1: Payoff at the symmetric stage game Prisoner's Dilemma ($A > 3.15$).

biological process - types are genetically determined, and the payoff is the expected number of offspring, and (2) learning and imitation process - an agent's type describes the way he perceives strategic interactions; once in a while an agent may decide to change his strategic framework and imitate another person's type, if the other person is more successful.

At each generation the agents in the population are randomly matched and each couple plays M times (without rematching) the symmetric stage game of the Prisoner's Dilemma with the payoffs given in Table 1:[2] mutual cooperation (both players play C) yields both players $A > 3.15$, mutual defection (both players play D) gives 1, and if a single player defects, he obtains $A + 1$ and his opponent gets 0. Note that the parameter A is the ratio between what can be gained by mutual cooperation to the additional payoff that is obtained by defecting.[3]

Each agent in our model has a type (level) in the set $\{L_1, ..., L_M\}$ that determines how many steps he looks ahead. An agent of type L_k looks k steps ahead in his strategic reasoning. When the *horizon* (the number of remaining stages) is larger than k the agent must follow a simple heuristic. We assume that this heuristic must satisfy two properties: (1) "nice" (never be the first player to defect), and (2) "retaliating" - defect if the opponent defected in the previous stage. Two examples for such heuristics are "grim" and "tit-for-tat".[4] When the horizon is equal to k, the agent begins to play strategically and he may choose any action. We interpret L_k's behavior to stem from bounded forward-looking: when the horizon is larger than k, he subjectively perceives it to be infinite, and he does not take into account the fact that the interaction has a well-defined final period, and that this final period has strategic implications. One can also consider our model as a reduced-form for an interaction with a random unknown long length, in which each type L_k gets a signal about the interaction's

[2] All our results are independent of the value of M (given that $M \geq 4$). The inequality $A > 3.15$ is required for the solution we characterize below to be evolutionary stable in a non-empty interval of p-s.

[3] We assume that defection yields the same additional payoff (relative to cooperation) regardless of the opponent's strategy to simplify the presentation of the result (but the results remain qualitatively similar also without this assumption). Given this assumption we normalize, without loss of generality, the payoff of being a single cooperator to be 0, and the additional payoff of defecting to be 1.

[4] *Grim* heuristic defects if and only if the opponent ever defected in the past, and *Tit-for-tat* heuristic defects if and only if the opponent defected in the previous stage. In Section 7 we discuss the extension of our model to a setup in which a player may choose his heuristic for long horizons, and the relation to the notion of analogy-based expectation equilibrium (Jehiel (2005)).

realized length k periods before the end (see Section 7). Note that the set of strategies of type L_k is a strict subset of the set of strategies of type L_{k+1}, and that type L_M is fully-rational and has an unlimited set of strategies.

We assume that types are partially observable in the following way (similar to Dekel, Ely, and Yilankaya (2007)): before the interaction begins, each agent has an independent probability p to observe his opponent's type.[5] Informally, this can be interpreted as an opportunity to observe your opponent's past behavior, or to observe a trait that is correlated with cognitive level (such as I.Q. level, see Gill and Prowse (2012)). The total payoff of an agent of type L_k is the undiscounted sum of payoffs in the repeated prisoner dilemma minus an arbitrarily small cost that is increasing in k (a marginal cost for having a better forward looking-ability).

In common with much of the evolutionary literature, we use a static solution concept to tractably capture the stable points of a dynamic evolutionary process. Specifically, we adapt the notion of evolutionary stable strategy (ESS, Maynard-Smith (1974)) to a setup with different types. In such a setup, the state of the population is described by a *configuration* (Dekel, Ely, and Yilankaya (2007)) - a pair consisting of a distribution of types and the (possibly mixed) strategy that each type uses in the game. A configuration is *evolutionary stable* if any sufficiently small group of mutants who invades the population is outperformed by the incumbents in the post-entry population.[6]

Evolutionary stability can be sustained by playing very badly when facing types outside the support of the distribution. However, this is unlikely to be stable in the long run, as the strategy played against a non-existing type should slowly evolve as a response to recurrent entrees of mutants. Thus, we refine neutral stability by requiring agents to play undominated strategies against non-existing types. In Section 6 we show that our results are robust to various plausible changes in the definition of stability in this setup, and to using the alternative notion of Dekel, Ely, and Yilankaya (2007).

Our main result shows that if p is not too close to 0 and 1 (and this interval is increasing in A), then there exists a unique evolutionary stable configuration with undominated strategies, which includes two kind of players: (1) *naive* agents of type L_1 who only begin defecting at the last stage, (2) *moderately-sophisticated* agents of type L_3: usually they defect two stages before the end, unless they observe that their opponent is sophisticated, and in this case, they begin defecting one stage earlier. The stability relies on the balance between the

[5] The results remain the same also in the case in which agents can only observe lower opponents' type (see section 6).

[6] The "mutants" achieve the same payoff if they are *equivalent* to the incumbents: have the same distribution of types and play the same on-equilibrium-path. If they are not equivalent, we require the mutants to achieve a strictly lower payoff.

direct disadvantage of naive agents - they defect too late, and the indirect "commitment" advantage - when naivety is being observed, it induces moderately-sophisticated opponents to postpone their defection (as naive agents are committed to cooperate longer), and this allows an additional round of mutual cooperation. The proportion of the naive players is increasing in both p and A.

It is interesting to note that stable configurations are very different when p is close to 0 or 1. In both cases, stable configurations must include fully-rational players who, when facing other fully-rational agents, defect at all stages. When p is close to 0, types are too rarely observed, and the indirect advantage of naive agents is too weak. When p is close to 1, there is an "arms-race" between sophisticated agents who observe each other: each such agent wishes to defect one stage before his opponent. The result of this "race" is that there must be some fully-rational agents in the population.

Existing evolutionary models that studied bounded strategic reasoning (Stahl (1993); Stennek (2000)) focused on the case where types are unobservable ($p = 0$) , and showed that in various games: (1) the most sophisticated type always survives, and (2) lower (more naive) types can also survive if they do not play serially dominated strategies. Recently, Mohlin (2012) showed that there may be evolutionary stable configurations in which the highest type do not survive, and he also studied the case in which higher types can perfectly observe lower types (a case similar to $p = 1$, see Section 6).[7] This paper focuses in a specific game (repeated prisoner dilemma) and allow partial observability (p strictly between 0 and 1), and this allow us to obtain a sharp and qualitative different prediction: only naive and moderately-sophisticated agents survive.

Existing experimental results verify the plausibility of both our assumption of using "nice" and "retaliating" heuristic for large horizons, and of our main prediction. Selten and Stoecker (1986) study the behavior of players in iterated Prisoner Dilemma games of 10 rounds (similar results are presented in Andreoni and Miller (1993); Cooper, DeJong, Forsythe, and Ross (1996); Bruttel, Güth, and Kamecke (2012)). They show that: (1) if any player defected, then almost always both players defect at all remaining stages, (2) usually there is mutual cooperation in the first 6 rounds, and (3) players begin defecting at the last 1-4 rounds.[8] Such behavior has two main explanations in the literature: (1) some players are altruistic, and (2)

[7] See also Crawford (2003) for a strategic (non-evolutionary) model of zero-sum games with "cheap-talk" in which naive and sophisticated agents may co-exist and obtain the same payoff.

[8] In Selten and Stoecker (1986)'s experiments players engaged in 25 sequences ("super-games") of iterated Prisoner's Dilemma. The above results describe the behavior of subjects in the last 13 sequences (after the initial 12 sequences in which players are inexperienced and their actions are "noisier"). During these 13 sequences there is a slow drift in the behavior of players towards earlier defections. Nevertheless, defections before the last 4 rounds were infrequent also in the last couple of rounds.

players have limited forward-looking.[9] Johnson, Camerer, Sen, and Rymon (2002) studied the relative importance of these explanations in a related sequential bargaining game, and their findings suggest the limited forward-looking is the main cause for this behavior.

A recent qualitative support for our prediction is given in Hyndman, Terracol, and Vaksmann (2012), which experimentally studied the strategic behavior of people across different games. They showed that a fraction of the players consistently assign a low level of reasoning to their opponent, while the remaining players alternate between different assessments of their opponent's cognitive skills. The former fraction corresponds to the "naive" agents in our model who always best-reply to a belief that the opponent is non-strategic and follows a "nice and retaliating" heuristic. The remaining players correspond to the "moderately-sophisticated" agents in our model who, depending on the signal they obtain, best reply to different beliefs about the opponent's cognitive skill.

The paper is structured as follows. Section 2 presents our model. In Section 4 we present our results, and it is followed by sketches of the proofs in Section **??** (formal proofs appear in the appendix). Sections 6 shows that our results are robust to various changes in the model. We conclude in Section 7.

2 Model

2.1 Strategies and Types

We study a symmetric finitely-iterated Prisoner's Dilemma game that repeats M stages ($M \geq 4$), denoted by G. The payoff of each stage game are as described in Table 1 ($A > 2 + \sqrt{2}$). This payoff is interpreted, as standard in the evolutionary literature, as representing "success" or "fitness". Define the horizon of a stage as the number of remaining stages including the current stage. That is, the horizon at stage m is equal to $M - k + 1$. History h_k of length k is a sequence of k pairs, where the l-th pair describes the actions chosen by the players at stage l. Let H_k be the sets of histories of length k, and let $H = \cup_{1 \leq k < M} H_k$ be the set of all non-terminal histories.

A pure strategy s is a function from H into $\{C, D\}$. A , and a behavioral strategy σ is a function from H into $\Delta(\{C, D\})$. With some abuse of notations we write $\sigma(h_k) = C$ when σ assigns probability 1 to playing C (and similarly for D). Let Σ be the set of behavioral strategies (henceforth, strategies). Strategy σ is *k-nice-retaliating* if whenever the horizon is larger than k: (1) σ assigns probability 1 to C if the opponent has never defected before, and

[9] Heifetz and Pauzner (2005) explain this behavior with a different kind of cognitive limitations: at each node, each player has a small probability to be "confused" and choose a different action than the optimal one.

(2) σ assigns probability 1 to D if the opponent has defected in the previous stage. Let Σ_k be the set of *k-Nice-Retaliating* behavioral strategies. Let $d_k \in \Sigma_k$ be the pure strategy that plays *grim* as long as the horizon is larger then k: defects if and only if the opponent has defected in the past, and then defects at all following stages (when the horizon is at most k). Let $\mathcal{D} = \{d_k\}_{0 \leq k \leq M}$ be the set of all such "grim-then-defect" strategies. Let $u(\sigma, \sigma')$ be the expected payoff of a player who plays strategy σ against an opponent who plays behavioral strategy σ'.

We imagine a large population randomly matched to play G. Different agents in the population differ in their cognitive ability, which is captured by their type. Let $\mathcal{L} = \{L_1, ..., L_M\}$ be the set of types (or levels).[10] An agent of type L_k looks only k steps ahead, and when the horizon is larger than k he ignores end-of-game strategic considerations and plays a "nice and retaliating" heuristic . That is, an agent with type L_k can only play *k-nice-retaliating* strategies. When the horizon is at most k, the agent is no longer limited in his play.

Let $c : \mathcal{L} \to \mathbb{R}^+$ a strictly increasing function satisfying $c(L_1) = 0$, and let $\delta > 0$. Agents of type L_k bear a *cognitive cost* of $\delta \cdot c(L_k)$. In the analysis in the following sections we will focus on the case where δ is sufficiently small (arbitrarily low cognitive costs). The payoff of the repeated game is the undiscounted sum of the stage payoffs minus the cognitive cost.

Following the model of partial observability of Dekel, Ely, and Yilankaya (2007), we assume that each player knows the type of his opponent with probability p (and get no information about his opponent's type with probability $1 - p$), independently of the event that his opponent knows his type. We use the term *stranger* to describe an opponent that his type was not observed. In Section 6 we demonstrate that our results remain the same also if agents can only identify their opponent's type if that type is lower.

2.2 Configurations

The state of the population is described by a *configuration* - a pair consisting of a distribution of types and the strategy that each type uses in the game. Formally (where $C(\mu)$denote the support of μ):

Definition 1. *Configuration (or population)* (μ, b) is a pair where $\mu \in \Delta(\mathcal{L})$ is the distribution of types in the population, and $b = (b_k)_{k \in C(\mu)}$ is the profile of signal-dependent strategies is played by each type in the population. That is, for each type $L_k \in C(\mu)$, $b_k : \mathcal{L} \cup \emptyset \to \Sigma_k$ is a signal-dependent strategy that specifies a behavioral *k-nice-retaliating* strategy for each

[10] We explicitly omit level 0 (L_0, who uses a nice and retaliating heuristic throughout the entire interaction). The results are qualitatively the same if L_0 is included (see Subsection 6.1.1).

possible observation about the opponent's type (including observations with zero probability of types outside $C(\mu)$).

Remark 1. We note two points regarding Definition 1:

- Agents of type L_k can use a behavioral (non-pure) strategy. As usual in such models, this can be interpreted as either: (1) each agent randomly chooses his actions, or (2) different fractions of type L_k play different pure strategies, and the aggregate distribution induces the randomness.

- A configuration also determines the strategies that are used against non-existing types ("mutant" types outside $C(\mu)$). In Section 6 we propose an alternative stability notion, according to which, a configuration only determines that strategies that are used against types with positive frequency (similar to the definition of a configuration in Dekel, Ely, and Yilankaya (2007)).

Given a configuration (μ, b), we call the types in $C(\mu)$ as *existing types* or *incumbents*, and types outside $C(\mu)$ as *non-existing* types or *mutant* types. Next, we define the mixture of two configurations as follows:

Definition 2. Let (μ, b) and (μ', b') be configurations, and let $0 < \epsilon < 1$. The *mixture configuration* $\left(\tilde{\mu}, \tilde{b}\right) = (1-\epsilon)(\mu, b) + \epsilon(\mu', b')$ is:

- $\tilde{\mu} = (1-\epsilon)\mu + \epsilon\mu'$.

- For each $k \in C(\tilde{\mu})$:
$$\tilde{b}_k = \frac{(1-\epsilon)\cdot\mu(L_k)\, b_k + \epsilon\cdot\mu'(L_k)\, b'_k}{\mu(L_k) + \mu'(L_k)}.$$

When ϵ is small we interpret $(1-\epsilon)(\mu, b) + \epsilon(\mu', b')$ as the *post-entry configuration :* a population of *incumbents* in state (μ, b) that was invaded by ϵ mutants with configuration(μ', b').

Finally, we define that two configurations are equivalent if they have the same distribution, and they induce the same observed play. Formally:

Definition 3. Configurations (μ, b) and (μ', b') are equivalent (denoted by $(\mu, b) \approx (\mu', b')$) if:

1. $\mu = \mu'$.

2. For each pair of types $L_k, L_{k'} \in C(\mu)$, the observed play when type L_k plays against type $L_{k'}$ is the same in both populations.

Note that two equivalent configurations induce the same observable play only on the equilibrium path. Following the invasion of ϵ mutants, the incumbents in each of the equivalent configurations may act very differently when facing mutants.

3 Evolutionary Stability

3.1 Solution Concept

In a model without types, the state of the population is described by a strategy. A strategy is neutrally (resp., evolutionary) stable if any sufficiently small group of mutants who invades the population and play an arbitrary strategy would achieve a weakly (strictly) lower payoff than the incumbents. Formally:

Definition 4. (Maynard-Smith (1974); Maynard Smith (1982)) Strategy $\sigma \in \Sigma$ is neutrally (resp., evolutionary) stable if for any "mutant" strategy σ' (resp., $\sigma' \neq \sigma$) there exists some $\epsilon_{\sigma'} \in (0,1)$ such that for every $0 < \epsilon < \epsilon_{\sigma'}$:

$$u\left(\sigma, \epsilon\sigma' + (1-\epsilon)\,\sigma\right) \geq u\left(\sigma', \epsilon\sigma' + (1-\epsilon)\,\sigma\right).$$

(resp., $u\left(\sigma, \epsilon\sigma' + (1-\epsilon)\,\sigma\right) > u\left(\sigma', \epsilon\sigma' + (1-\epsilon)\,\sigma\right)$).

In what follows we extend the notion of evolutionary stability from strategies to configurations. Given two configurations (μ, b) and (μ', b') define $u\left((\mu, b), (\mu', b')\right)$ as the expected payoff of a player from population (μ, b) who plays against an opponent from population (μ', b') (and the type of each player is observed with independent probability p). A configuration is *neutrally (evolutionary) stable* if any sufficiently small group of mutants who invades the population would obtain a weakly (strictly) lower payoff than the incumbents in the post-entry population. Formally:

Definition 5. Configuration (μ, b) is *neutrally* (resp., *evolutionary*) *stable* if for any "mutant" configuration (μ', b') (resp., any $(\mu', b') \not\approx (\mu, b)$) there exists some $\epsilon_{\sigma'} \in (0,1)$ such that for every $0 < \epsilon < \epsilon_{\sigma'}$:

$$u\left((\mu, b), \epsilon\,(\mu', b') + (1-\epsilon)\,(\mu, b)\right) \geq u\left((\mu', b'), \epsilon\,(\mu', b') + (1-\epsilon)\,(\mu, b)\right).$$

(resp., $u\left((\mu, b), \epsilon\,(\mu', b') + (1-\epsilon)\,(\mu, b)\right) > u\left((\mu', b'), \epsilon\,(\mu', b') + (1-\epsilon)\,(\mu, b)\right)$.)

Definition 5 is closely related to Maynard Smith (1982)'s Definition 4 in two ways:

1. When the set of types is a singleton, then Definition 5 and Definition 4 coincide.

2. Consider the following 2-player "meta-game": each player chooses a type L_k and a signal-dependent *k-nice-retaliating* strategy. Note that a mixed strategy in this meta-game is a configuration, and a neutrally stable strategy in the "meta-game" is a neutrally stable configuration.

Remark 2. Note that:

1. Any evolutionary stable configuration is also neutrally stable.

2. Evolutionary stable configurations are only weakly stable to invasions of mutants who play exactly like the incumbents on-equilibrium-path.

With some abuse of notation we denote by L_k also the distribution that assigns mass 1 to type L_k. It is well-known that any neutrally stable strategy is a Nash equilibrium. Proposition 1 shows that the strategy profile in a neutrally stable configuration is: (1) *balanced* - yield the same payoff to all types in the population, and (2) Bayesian-Nash equilibrium.

Proposition 1. *Let (μ, b) be a neutrally stable configuration. Then, the strategy profile b: (1) induces the same payoff for each type in the support of μ, and (2) is a Bayes-Nash equilibrium in the Bayesian game with distribution of types μ.*

Proof.

1. Assume to the contrary that b induces different payoffs to different types. Let $L_k \in C(\mu)$ be the type with the highest payoff. Then: $u((L_k, b_k), (\mu, b)) > u((\mu, b), (\mu, b))$. This implies that for sufficiently small $\epsilon > 0$, mutants of type L_k who play b_k would achieve a strictly higher payoff than the incumbents and this contradicts the stability of (μ, b).

2. Assume to the contrary that b is not a Bayesian-Nash equilibrium. Let $L_k \in C(\mu)$ be the type who does not play a best response against (μ, b). This implies that there exists strategy b'_k such that : $u((L_k, b'_k), (\mu, b)) > u((L_k, b_k), (\mu, b))$. By the first part of the proposition, $u((L_k, b_k), (\mu, b)) = u((\mu, b), (\mu, b))$. This implies that for sufficiently small $\epsilon > 0$, mutants of type L_k who play b'_k would obtain a strictly higher than the incumbents and this contradicts the stability of (μ, b).

□

3.2 Result (Stability)

Our first result characterizes an evolutionary stable configuration, (μ^*, b^*) , in which naive players (type L_1) and moderately-sophisticated players (type L_3) co-exist. Let the configuration (μ^*, b^*) be defined as follows:

1. The population includes only types L_1 and L_3:

$$\mu^*(L_1) = \frac{p(A-1) - 1 + \delta \cdot c(L_3)}{p(A-1)}, \quad \mu^*(L_3) = \frac{1 - \delta \cdot c(L_3)}{p(A-1)}.$$

2. The "naive" Agents of type L_1 play d_1 : use "grim" until the last stage, and defect at the last stage.

3. The "moderately-sophisticated" agents of type L_3 play:

   (a) d_2 against strangers and observed L_1 (follow "grim" until the last 2 stages, and defect in last 2 remaining stages).

   (b) d_3 against any observed type different then L_1.

Theorem 1. *Let* $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$ *and let* $\delta > 0$ *be sufficiently small. Then* (μ^*, b^*) *is evolutionary stable.*

The formal proof appears in Appendix A.1. In what follows we briefly sketch the outline of the proof. First, we show that b^* is a Bayes-Nash equilibrium (given μ^*). Naive players (L_1) play their unique dominating strategy - d_1 (as they must follow their "nice and retaliating" heuristic when the horizon is larger than 1). For sophisticated players, playing d_3 against sophisticated opponents is strictly better than an earlier defection for small enough p, and playing d_2 against strangers and naive opponents is strictly better than earlier defections if $\mu(L_1)$ is large enough.

Next, we show that (μ^*, b^*) is balanced. In order to show it, we compare the fitness of naive and sophisticated agents as a function of their opponent. Naive agents succeed more only against an observing sophisticated opponent (who observed their type), because their observed naivety induces an additional round of mutual cooperation. Sophisticated agents achieve a better payoff in the two other cases: against naive opponents and against an unobserving sophisticated opponent. This implies that there is a unique level of $\mu(L_1)$ that balances the payoff of the two kinds of players. Finally, we use these two properties to show resistance to mutations. If ϵ more naive players join the populations, then due to the previous arguments, naive agents would have a strictly lower payoff than the incumbents (on average). The same holds for ϵ more moderately-sophisticated who join the population.

4 Uniqueness

4.1 Undominated Configurations

The interaction admits additional evolutionary stable configurations. One such configuration is described in the following example.

Example. Consider the configuration that assigns mass 1 to fully-rational agents (type L_M) who deviate at all stages against any observed opponent's type. One can see that this configuration is evolutionary stable. However, the stability relies on the incumbents defecting at all stages against naive mutants (L_1). Such a strategy is strictly dominated by an alternative strategy that cooperates for the first $M-2$ stages against naive opponents. Thus, in the long run, as a response to recurrent entrees of naive mutants, incumbents are expected to evolve into cooperation at the first stages of the game when facing naive opponents, and the stability of the configuration will be lost.

Motivated by this example, we refine neutral stability by not allowing agents to use "bad" strategies against non-existing types. The payoff of an incumbent's strategy that is played against a mutant type, depends on that mutant's strategy. One may expect that most of the time invading mutants will best-reply to the incumbents because (see, Swinkels (1992)) either: (1) "best-reply" mutants have higher fitness than other mutants, and thus they are expected to survive longer in the post-entree population; or/and (2) mutants choose their strategy by experimentation, and they are more likely to choose best-reply strategies. Formally:

Definition 6. Let (μ, b) be a configuration and let $L_{\tilde{k}} \in \mathcal{L} \backslash C(\mu)$ be a mutant type. A signal-dependent strategy $\tilde{b}_{\tilde{k}} : L \cup \emptyset \to \Sigma_{\tilde{k}}$ is *best-reply* if $u\left(\left(L_{\tilde{k}}, \tilde{b}_{\tilde{k}}\right), (\mu, b)\right) \geq u\left(\left(L_{\tilde{k}}, b'_{\tilde{k}}\right), (\mu, b)\right)$ for each alternative signal-dependent strategy $b'_{\tilde{k}} : L \cup \emptyset \to \Sigma_{\tilde{k}}$.

An incumbent strategy $b_k\left(\tilde{k}\right)$ is dominated by another strategy $b'_k\left(\tilde{k}\right)$ if it yields a strictly worse payoff against all best-reply of mutants of type $L_{\tilde{k}}$. Strategy $b_k\left(\tilde{k}\right)$ is undominated if it is not dominated by any other strategy. Formally:

Definition 7. Let (μ, b) be a configuration, let $L_k \in C(\mu)$ be an incumbent type, let $L_{\tilde{k}} \in \mathcal{L} \backslash C(\mu)$ be a mutant type, let $b'_k\left(\tilde{k}\right) \in \Sigma_k$ be strategy, and for each $k' \neq \tilde{k}$ let $b'_k(k') = b_k(k')$. Strategy $b_k\left(\tilde{k}\right) \in \Sigma_k$ is *dominated by* $b'_k\left(\tilde{k}\right)$ if for each best-reply signal-dependent strategy $\tilde{b}_{\tilde{k}} : L \cup \emptyset \to \Sigma_{\tilde{k}}$: $u\left(\left(L_k, b_k\right), \left(L_{\tilde{k}}, b_{\tilde{k}}\right)\right) < u\left(\left(L_k, b'_k\right), \left(L_{\tilde{k}}, b_{\tilde{k}}\right)\right)$. Strategy $b_k\left(\tilde{k}\right) \in \Sigma_k$ is *undominated* if there does not exist strategy $b'_k\left(\tilde{k}\right) \in \Sigma_k$ such that $b_k\left(\tilde{k}\right)$ is dominated by $b'_k\left(\tilde{k}\right)$.

We refine the notion of neutral stability by requiring all strategies that are played against non-existing mutants to be undominated. Formally:

Definition 8. Configuration (μ, b) is *undominated*, if for each "incumbent" type $L_k \in C(\mu)$, and for each "mutant" type $L_{\tilde{k}} \in \mathcal{L} \backslash C(\mu)$, the strategy $b_k\left(\tilde{k}\right)$ is undominated.

A Configuration is *undominated neutrally (evolutionary) stable* if it is both undominated and neutrally (evolutionary) stable.

4.2 Result (Uniqueness)

Our second shows that any undominated neutrally stable configuration must be equivalent to (μ^*, b^*).

result gives a sharp prediction for the unique undominated neutrally stable configurations in the interval $\frac{A}{(A-1)^2} < p < \frac{A-2}{A-1}$. In this configuration naive players (type L_1) and moderately-sophisticated players (type L_3) co-exist. Formally:

Theorem 2. *Let $\frac{1}{A-1} < p < 1 - \frac{2 \cdot A-1}{A^2-A}$, and let (μ, b) be an undominated neutrally stable configuration. Then (μ, b) and (μ^*, b^*) are equivalent.*

The sketch of the proof is as follows (the formal proof is given in Appendix A.2). First, observe that a configuration with a single type is not stable: 1) if the type is L_M, then the entire population defects all the time, and mutants of type L_1 would induce cooperation against them and invade the population; and 2) if the type is $L_k \neq L_M$, then mutants of type L_{k+1} can invade the population and get strictly higher payoff then the incumbents. Let L_k be the smallest ("naive") type in the population. Then, it is immediate to see that type L_k must always defect when the horizon is at most k (as it is common knowledge that all players are rational at that stage), and all other types must defect when the horizon is at most $k+1$.

The next step is to show that a large fraction of the non-naive population must cooperate at all horizons larger than $k+1$ when facing strangers. Otherwise, a small increase in the frequency of the naive players, would improve their fitness relative to the non-naive agents (as many non-naive loose rounds of mutual cooperation while defecting earlier then $k+1$ against strangers), and the configuration will be unstable. The fact that this fraction is so large, implies that if there are non-naive players who defect at earlier horizon against strangers, then: (1) the large fraction who defects at horizon $k+1$ against strangers must belong to type L_{k+1}, and (2) all the remaining players (type larger than $k+1$) must defect at horizon $k+2$ against strangers and one stage before an observed opponent (who has not observed their

type). This characterization allows to find the unique distribution of types who satisfy the balance of payoffs, but it turns out that this distribution is not stable to small perturbations in the frequency of the different types.

Finally, if all non-naive players defect at horizon $k+1$ against strangers, then it implies that they all defect at horizon $k+2$ against observed non-naive opponents, and the balance between the payoffs of the different types imply that the frequency of naive and non-naive players is like in μ^*. Finally, we show that if $k > 1$, then the configuration can be invaded by mutants of type L_1, who would earn from inducing more mutual cooperation when being observed by their opponent.

5 Stability for Low and High p-s

Our next result, shows that in the benchmark cases when p is close to 0 and 1 the undominated neutrally stable configurations are very different. In both cases, stable configurations must include fully-rational players who, when facing other fully-rational agents, defect at all stages. When p is close to 0, this occurs because the indirect advantage of lower types is too small and they can not exist in a stable configuration (because the probability of being identified by the opponent is too low). When p is close to 1, there is an "arms-race" between sophisticated agents who observe each other: each such agent wishes to defect one stage before his opponent. The result of this "race" is that there must some fully-rational agents in the population. Formally:

Theorem 3.

1. *Let $0 \leq p < \frac{1}{(M-2)\cdot(A-1)}$. Then there exists an undominated evolutionary stable configuration $\left(\tilde{\mu}, \tilde{b}\right)$ where all players have type L_M and they play d_M against strangers and type L_M, and d_{k+1} against observed "mutant" type $L_k < L_M$. Moreover, any other undominated neutrally stable configuration is equivalent to $\left(\tilde{\mu}, \tilde{b}\right)$.*

2. *Let $1 \geq p > \frac{A-1}{A}$. Then in any undominated neutrally stable configuration there is a positive frequency of players of type L_M, and these players defect at all stages when observing an opponent of type L_M.*

The sketch of the proof is as follows (formal proof appear in Appendix A.3):

1. **Low p-s:** The configuration that everyone has type L_M (fully-rational) and begin defecting at the first stage is stable because the indirect advantage of naive mutants (with a lower type than L_M) is too small: they strictly lose when their naivety is

unobserved, and their naivety is observed too rarely. Due to a similar argument, in any other configuration where different types co-exist, the lower type would obtain a strictly lower payoff (and this implies the uniqueness).

2. **High p-s:** Assume to the contrary that no agent in the population ever defects at the first stage. Let $l < M$ be the horizon in which the highest type in the population begin defecting when they observe an opponent of the same type. If p is large enough, then their opponent is likely to observe their signal as well and begin defecting at stage l as well. This implies (again for large enough p) that starting to defect one stage earlier is strictly better. This implies that mutants who "imitate" the highest type's behavior except defecting one stage earlier when observing an opponent with the highest type, would achieve a strictly higher payoff.

6 Robustness

In this section we demonstrate that our results are robust to various plausible changes in the model. In Subsection 6.1 we deal with variants in the types and in the signal structure, and in Subsection 6.2 we deal with different stability notions.

6.1 Variants in the Model

6.1.1 Level 0

In the model we do not allow players to belong to "level-0" (L_0) who follow a *nice and retaliating* strategy at all rounds of the interaction. Such "level-0" players play a strictly-dominated strategy (cooperating at the last stage), and we chose to omit them from the model as such extreme bounded forward-looking may seem implausible. We note that our results are qualitatively robust to the addition of type L_0 in the following sense. All of our results would remain shift a single step backwards: the naive players in the stable configurations in $\mathcal{C}$ would be of type L_0 instead of L_1, and the sophisticated players would look 1-2 steps ahead instead of 2-3 steps.

6.1.2 Asymmetric Type Observability

In the model we assume that any agent has the same probability to observe his opponent's type. In particular, lower types may identify the exact type of a more sophisticated opponent. One may argue (see, e.g., Mohlin (2012)) that it is more plausible that only higher types can identify the type of their opponents. We formalize this alternative assumption as follows.

Before the interaction begins each agent independently obtains a signal about his opponent. With probability $1-p$ the signal is non-informative ($\emptyset$). With probability p the signal is informative:

1. If the opponent's type is strictly lower, then the agent exactly identifies it.
2. If the opponent's type is weakly higher, then the agent only observes that is opponent's type is weakly higher than his own type.

One can see that all of our results remain the same in this setup.

6.1.3 Small Perturbations to the Signal Structure

Our results remain qualitatively similar if the signal structure is slightly altered by any of the following perturbations:

1. There is a small positive correlation between the signal that each agent obtains about his opponent's type.
2. There is a small chance that the informative signal is incorrect.

That is, if the perturbation is small enough, then there exists a unique undominated neutrally stable configuration which is closed to (μ^*, b^*)).

6.2 Different Stability Notions

6.2.1 Focal Stability

One may argue that it is more plausible that the state of the population only specifies the behavior of players against existing types, and the behavior against mutant that introduce new types should be evolve as part of a post-entry adaptation process. In what follows we formalize this idea, and present an alternative notion of focal stability, and state that all our results remain the same with this stability notion (which may be of independent interest in future research).

A *compact configuration* is a pair consisting of a distribution of types and the strategy that each type uses against other types in the support of the distribution. Formally:

Definition 9. *Compact Configuration* (μ, b) is a pair where $\mu \in \Delta(\mathcal{L})$ is the distribution of types in the population, and $b = (b_k)_{k \in C(\mu)}$ is the profile of signal-dependent strategies is played by each type in the population given any signal with positive probability. That is, for each type $L_k \in C(\mu)$ in the population, $b_k : C(\mu) \cup \emptyset \to \Sigma_k$ is a signal-dependent strategy that

specifies a *k-nice-retaliating* strategy for each possible observation (with positive probability) about the opponent's type.

Given a compact configuration (μ, b), an invading *mutant configuration* (μ', b') should specify the signal-dependent *k'-nice-retaliating* strategy of each mutant type $L_{k'} \in C(\mu')$ against types in the support of the post-entry population $C(\mu) \cup C(\mu')$. *Internal mutant configurations* are those that do not introduce new types to the population: $C(\mu') \subseteq C(\mu)$. Internal mutants are interpreted as the combination of small perturbations to the frequency of incumbent types, and experimentation of new strategies by a small group in the population. A compact configuration is internally neutrally (evolutionary) stable if any sufficiently small group of (non-equivalent) internal mutants would obtain a weakly (strictly) lower payoff than the incumbents in the post-entry population. Formally:

Definition 10. Compact configuration (μ, b) is *internally neutrally (evolutionary) stable* if for any internal mutant configuration (μ', b') $((\mu', b') \not\approx (\mu, b))$ with $C(\mu') \subseteq C(\mu)$ there exists some $\epsilon_{\sigma'} \in (0, 1)$ such that for every $0 < \epsilon < \epsilon_{\sigma'}$:

$$u\left((\mu, b), \epsilon(\mu', b') + (1-\epsilon)(\mu, b)\right) \geq u\left((\mu', b'), \epsilon(\mu', b') + (1-\epsilon)(\mu, b)\right).$$

$(u\left((\mu, b), \epsilon(\mu', b') + (1-\epsilon)(\mu, b)\right) > u\left((\mu', b'), \epsilon(\mu', b') + (1-\epsilon)(\mu, b)\right).)$

External mutants are those that introduce a new type to the population. In this case, we assume that the incumbent population and the new mutant type interactively adapt their joint behavior, while taking the "focal" behavior of incumbents against other incumbents and strangers as fixed. We further assume that this adaptation process is fast enough relative to the evolution of types, such that the behavior in the post-entry population converge into a Bayesian-Nash equilibrium. A compact configuration is *(strictly) externally focally stable* if any mutant with a new type would achieve a (strictly) worse payoff in the induced post-entree Bayesian-Nash equilibrium. Formally:

Definition 11. Given a compact configuration (μ, b), $\epsilon > 0$ and an external mutant type $L_{k'} \in \mathcal{L} \backslash \mathcal{C}(\mu)$ let $B(\mu, b, L_{k'}, \epsilon)$ be the set of *post-entree adjusted configurations*: the set of configurations (μ', b') that satisfy:

1. The post-entry distribution is a mixture of ϵ mutants and $1-\epsilon$ incumbents: $\mu' = (1-\epsilon) \cdot \mu + \epsilon \cdot L_{k'}$.

2. The incumbents continue to play the same (focally) as in the pre-entry configuration against strangers and other incumbents: $b'_k(\emptyset) = b_k(\emptyset)$ for each $k \in C(\mu)$, (3) $b'_k\left(\tilde{k}\right) = b_k\left(\tilde{k}\right)$ for each $L_k, L_{\tilde{k}} \in C(\mu)$.

3. Each incumbent type best reply when facing an observed mutant.

4. The mutant type best replies to all opponents.

Definition 12. Compact configuration (μ, b) is *(strictly) externally focally stable* if for any mutant type $L_{k'} \in \mathcal{L} \backslash \mathcal{C}(\mu)$ there exists some $\epsilon_{\sigma'} \in (0,1)$ such that for every $0 < \epsilon < \epsilon_{\sigma'}$ and in any post-entree adjusted configuration $(\mu', b') \in B(\mu, b, k', \epsilon)$ the mutant obtains a lower payoff:

$$u((\mu, b'), (\mu', b')) \geq u((L_{k'}, b'), (\mu', b')).$$

$(u((\mu, b'), (\mu', b')) > u((L_{k'}, b'), (\mu', b')).)$

Finally, a compact configuration is (strictly) *focally stable* if it is both neutrally (evolutionary) stable and (strictly) externally focally stable. Simple adaptations to the proofs in the appendix yield the same results with focal stability. Formally (proof is omitted):

Proposition 2. *Let* $\frac{A}{(A-1)^2} < p < 1 - \frac{2 \cdot A - 1}{A^2 - A}$ *and let* $\delta > 0$ *be sufficiently small. Then the compact configuration* (μ^*, b^*) *is strictly focally stable. Moreover, if* (μ, b) *is a focally stable configuration then* (μ, b) *and* (μ^*, b^*) *are equivalent.*

6.2.2 DEY-Stability (Dekel, Ely, and Yilankaya (2007))

In our definition of focal stability the incumbents only approximately best-reply in the post-entry population, because they keep their play against incumbents and strangers the same, and do not adjust it to the presence of the new ϵ mutants. In some evolutionary setups, the adaptation process according to which agents choose their strategies might be much faster then the evolutionary process according to which the frequency of the types evolve. In these setups, it may be plausible to assume that the post-entree population adjust their play into an exact Bayesian-Nash equilibrium after any entree of mutants (both external and internal mutants).

Dekel, Ely, and Yilankaya (2007)'s notion of stability makes this assumption.[11] A compact configuration (μ, b) is *(strictly) DEY-stable* if:

1. The strategy profile b is:

   (a) A Bayesian-Nash equilibrium in the Bayesian game with the distribution of types μ.

[11] A similar approach is also used in the notions of *mental equilibrium* (Winter, Garcia-Jurado, and Mendez-Naya (2010)) and *evolutionary-stable types (Alger and Weibull (2012)).* Both notions apply only to homogenous populations that includes a single type, and thus are less appropriate to study stability of heterogeneous populations.

(b) Balanced - it induces the same payoff to all types in $C(\mu)$.

2. For each "mutant" type $L_k \in \mathcal{L}$, there exists sufficiently small ϵ_0 such that for each $\epsilon < \epsilon_0$, after ϵ mutants of type L_k invade the population:

   (a) There exist post-entry Bayesian-Nash equilibria in which the incumbents play is only slightly changed relative to the pre-entry play.

   (b) In all these equilibria the mutants would achieve a (strictly) lower payoff than the incumbents.

With simple adaptations, Lemmas 3-4 apply also for DEY-stability. This immediately implies the following theorem.

Proposition 3. *Let* $\frac{A}{(A-1)^2} < p < 1 - \frac{2 \cdot A - 1}{A^2 - A}$ *and let* $\delta > 0$ *be sufficiently small. Then the compact configuration* (μ^*, b^*) *is strictly DEY-stable.*

Moreover, any other DEY-stable configuration (μ, b) *has similar qualitative properties:*

1. *Naive agents of type* L_1 *exist in the population.*

2. *Moderately-sophisticated types (a non-empty subset of* $\{L_2, L_3, L_4\}$*) co-exist together with the naive type.*

3. *Higher levels of sophistication (*L_5*and above) do not exist.*

The reasons that we have to replace the uniqueness with the weaker "qualitative uniqueness" is that Lemma 6 does not apply for DEY-stability:

- Part (1) of Lemma 6 does not hold in this setup, because after ϵ mutants of type $L_{\tilde{k}}$ who always play d_3 enter the population, the incumbents adjust their strategies into an exact Bayesian-Nash equilibrium by lowering the frequency of incumbent of type $L_{\tilde{k}}$ would play d_3. We note that this adjustment that is implied by Dekel, Ely, and Yilankaya (2007)'s definition works in the opposite direction to the incentives that the incumbents face: if a random perturbation slightly increased the frequency of $L_{\tilde{k}}$ play d_3, then the incumbents who play d_3 obtain a higher payoff, but the adjustment process into a new equilibrium lowers their frequency in the population.

- Part (4) of Lemma 6 because Dekel, Ely, and Yilankaya (2007)'s definition only considers entry of mutants with a single type.

Finally, we note that if one adapts Dekel, Ely, and Yilankaya (2007)'s definition by assuming that the adjustment into a new exact equilibrium takes place only after the entree of external mutants, then all of our results, including the uniqueness and Lemma 6 would hold.

7 Concluding remarks

1. **Other heuristics for long horizons**: In our model we assumed that all players use *nice* and *retaliating* heuristics whenever the horizon is larger than their forward-looking ability. One could relax this assumption by allowing a player to choose his strategy for long horizons from some fixed set of heuristics. For example, the set of possible heuristics might be the strategies with "memory-1" (which depend only on the actions observed in the previous stage). Note that these "memory-1" strategies include the "grim" and "tit-for-tat" heuristics. A strategy of a player of type L_k in this setup specifies two strategic components for each possible signal about the opponent's type: (1) the heuristic he plays when the horizon is larger than k, and (2) the (unrestricted) strategy he plays when the horizon is at most k. It is immediate to apply our first result (Proposition 4) in this setup, and show that any configuration in $\mathcal{C}$ in which all players choose *grim* as their heuristic is stable. We conjuncture that there are only two sets of stable configurations in this extended setup: (1) efficient configurations: type distribution and strategies are equivalent to (μ^*, b^*), all players use heuristics that cooperate as long as the opponent never defected before, and a large enough proportion of each type defects if the opponent defected in the previous stage (a *nice* and *retaliating* heuristic); and (2) inefficient configurations in which all players defect at all stages (and use "always-defect" heuristic).

2. **Analogy-based expectation equilibrium:** Our model of bounded forward looking types could also be formulated using Jehiel (2005)'s Analogy-Based Expectation Equilibrium (ABEE). In this formulation a player of type L_k bundles all nodes with horizon of at least k into a single analogy class (while fully-differentiating among nodes with horizons smaller than k), and expects his opponent to play the same in all nodes of this class. The requirement that players play an evolutionary refinement of a Bayesian-Nash equilibrium in a configuration is replaced with the requirement that players play an analogous evolutionary refinement of ABEE in a configuration: at each stage every player best-responses to his analogy-based expectations, and expectations correctly represent the average behavior in every class. As in the previous remark: (1) it is immediate to show that every configuration in $\mathcal{C}$ is stable in this formulation (and players choose to play a *nice* and *retaliating* heuristic in their non-trivial analogy class), and (2) we conjuncture that there are only two sets of stable configurations in this ABEE formulation: efficient (μ^*, b^*)-like configurations, and inefficient "always-defect" configurations.

3. **Random continuation probability**: Our model assumes that the repeated interaction has a deterministic constant length, and that players completely ignore this fact when the horizon is too large. These assumptions may seem unrealistic. However, one should note that the model may be a reduced-form for a more realistic interaction with a random length and incomplete information. Specifiably, let $\mathbf{T}$ be the random unknown length of each interaction. Assume that the interaction lasts at least M rounds ($Pr(\mathbf{T} \geq M) = 1$), and that the continuation probability at each stage ($Pr(\mathbf{T} > n | \mathbf{T} > n-1)$) is not too far from 1. Bounded forward-looking is modeled in this setup as the stage in which a player becomes aware to the timing of the final period: player of type L_k gets a signal about the final period of the interaction (i.e., about the realization of $\mathbf{T}$) k stages before the end. In this setup, players are not restricted in their strategies (each type may play any strategy at any horizon). As in the previous remarks: (1) it is immediate to see that any configuration in (μ^*, b^*) is stable, and (2) we conjuncture that there are only two sets of stable configurations: efficient (μ^*, b^*)-like configurations, and inefficient "always-defect" configurations.

4. **Other games**: The formal analysis deals only with iterated Prisoner's Dilemma. However, we conjuncture that the results can be extended to other games in which iterated reasoning decreases payoffs. In particular, the extension of our results to "centipede"-like games (Rosenthal (1981)) is relatively straightforward. Such game can represent sequential interactions of gift exchanges. Such interactions were important in primitive hunter-gatherer populations (see, e.g., Haviland, Prins, and Walrath (2007), P. 440), which driven the biological evolution of human characteristics.

A Proofs

A.1 Stability of (μ^*, b^*)

Proposition 4. *Let* $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$ *and let* $\delta > 0$ *sufficiently small. Then the configuration* (μ^*, b^*) *(characterized in Theorem 2) is an undominated strongly neutrally stable.*

Proof. It is immediate to see that (μ^*, b^*) does not use strictly dominated strategies against mutant types and thus it is undominated configuration. In order to prove that is is neutrally stable, we first show two auxiliary results: (μ^*, b^*) is balanced (Lemma 1), and b^* is a Bayesian-Nash equilibrium (given μ^*) which is strict with respect to on-equilibrium-path deviations (Lemma 2).

Lemma 1. *Configuration* (μ^*, b^*) *is balanced (induce the same payoff to all types in* $C(\mu^*)$*).*

Proof. Let $q = \mu(L_1)$ be the frequency of the naive players. A naive player gets $(L-1)A+1$ against a naive opponent, and $(L-2)A+1$ against a sophisticated opponent (type L_3). A sophisticated player gets $(L-2)A+(A+1)+1 = (L-1)A+2$ against a naive opponent, and against a sophisticated opponent he gets: $(L-3)A+3$ if both players identify each other, $(L-3)A+(A+1)+2 = (L-2)A+3$ if only he identifies his opponent, $(L-3)A+0+2$ if only his opponent identifies him, and $(L-2)A+2$ if both players identify each other. Denote by $\delta_3 = \delta \cdot c(L_3)$ the cognitive cost of type L_3. The different types get the same payoff if:

$$q \cdot ((L-1)\cdot A+1) + (1-q)\cdot((L-2)\cdot A+1) + \delta_3 = q\cdot((L-1)\cdot A+2) + (1-q)\cdot \\ \left(p^2((L-3)A+3) + p(1-p)(((L-2)\cdot A+3) + ((L-3)\cdot A+2)) + (1-p)^2((L-2)\cdot A+2)\right)$$

$$(1-q)\left((L-2)A+1-\left((L-3)A+1+2p^2+p(1-p)(A+2+1)+(1-p)^2(A+1)\right)\right)+\delta_3 = q$$

$$q = (1-q)\cdot\left(A-\left(2p^2+p(1-p)(A+3)+(1-p)^2(A+1)\right)\right)+\delta_3$$

$$q = (1-q)\left(A-\left(p^2(2-A-3+A+1)+p(A+3-2A-2)+(A+1)\right)\right)+\delta_3$$

$$q = (1-q)(A-(p(1-A)+(A+1)))+\delta_3$$

$$q = (1-q)(-p(1-A)-1)+\delta_3 = (1-q)(p(A-1)-1)+\delta_3$$

$$q(p(A-1)-1+1) = p(A-1)-1+\delta_3$$

$$q = \frac{p(A-1)-1+\delta_3}{p(A-1)}. \tag{1}$$

Note that for each $p > \frac{1}{A-1}$ we get a valid value of $0 \le q \le 1$ for sufficiently small δ. □

Lemma 2. *The strategy profile b^* is a Bayesian-Nash equilibrium given a distribution of*

types μ^*. Moreover, any deviation that induces a different play on-equilibrium-path, yields a strictly worse outcome.

Proof. We have to show that both types play a best response (among the *k-nice-retaliating* strategies). This is immediate for a naive player (L_1), as his only choice is between cooperating and defecting at the last stage, and the latter strictly dominates the former. We have to show that a sophisticated player (L_3) play best response. It is immediate that d_2 is a strict best response against an observed naive opponent. Next, we show that playing d_2 against a stranger is strictly better than playing d_3. This is true if the following inequality holds (looking at the payoff of the last 3 rounds):

$$q(2A+2)+(1-q)(2p+(1-p)(A+2))>q(A+3)+(1-q)(3p+(1-p)(A+3))$$

$$q(A-1)>(1-q) \Leftrightarrow q>\frac{1}{A}.$$

Using (1) one obtains:

$$\frac{p(A-1)-1}{p(A-1)}>\frac{1}{A} \Leftrightarrow pA(A-1)-A>p(A-1)$$

$$pA^2-pA-A>pA-p \Leftrightarrow p\left(A^2-2A+1\right)>A \Leftrightarrow p>\frac{A}{(A-1)^2}.$$

It is immediate that d_2 is also strictly better (against strangers) than any other strategy that induces a different play on-equilibrium-path. We are left with showing that it is strict better for a sophisticated player to play d_3 and not d_4 against a sophisticated opponent (and this immediately implies that d_3 is strictly better against identified sophisticated opponents than any other strategy that induces a different play on-equilibrium-path). This is true if the following inequality holds (focusing on the payoffs of the last 4 rounds, as all preceding payoffs are the same):

$$p(A+3)+(1-p)(2A+3)>p(A+4)+(1-p)(A+4)$$

$$(1-p)(A-1)>p \Leftrightarrow A-1>Ap \Leftrightarrow p<\frac{A-1}{A}.$$

□

We now use the lemmas to prove that (μ^*, b^*) is strongly neutrally stable. That is, we have to show that after an invasion of ϵ mutants of configuration (μ, b) ($(\mu, b) \not\approx (\mu^*, b^*)$), the

incumbents obtain a strictly higher payoff than the mutants in the post-entree population (for sufficiently small $\epsilon > 0$).

Consider first mutants of types L_1 or L_3 (which exist in the pre-entry population). If these mutants play differently against incumbents (strangers, L_1 or L_3) then they earn strictly worse by the previous lemmas. Thus, such mutants must play the same against incumbent types and strangers (and may only differ in their play against "mutant" types other than L_1 and L_3). Denote such mutants as "imitating" mutants. Note that when the proportion of naive agents become larger (smaller) relative to its proportion in μ^* (and agents still follow b^*), then the naive agents achieve a lower (higher) payoff than the sophisticated agents. This is because naive agents obtain a strictly lower payoff than sophisticated agents, when facing naive opponents (the sophisticated players obtain an additional fitness point by defecting when the horizon is equal to 2). This implies that "imitating" mutants obtain a strictly lower payoff than the incumbents when facing incumbents or "imitating" mutants (unless the "imitating" mutants have the same distribution of types as the incumbents, and then they obtain the same payoff).

Next, consider mutant of different types (L_2 or L_4 or more). Mutants of type L_2 achieve a strictly lower payoff against incumbents: they have the same payoff as L_3 in most cases, but they obtain a strictly lower payoff when they observe an opponent of type L_3 (due to their inability to defect 3 stages before the end). Mutants of higher types (L_4or more) obtain at most the incumbents' payoff when facing incumbents, while they have a strictly larger cognitive cost ($\delta \cdot c(L_4)$). Thus also these mutants achieve a strictly lower payoff than the incumbents. Finally, mutants may gain an advantage from a *secret-handshake* like behavior () - playing the same against incumbent types and strangers, while cooperating with each other when observing a mutant type (different then L_1 and L_3). However, for sufficiently small ϵ, such an advantage cannot compensate for the strict disadvantages mentioned above, and this implies that any configuration of mutants would obtain a strictly worse payoff than the incumbents (unless they are equivalent, and then they obtain the same payoff). □

A.2 Uniqueness of (μ^*, b^*)

Proposition 5. *Let* $\frac{1}{A-1} < p < 1 - \frac{2 \cdot A - 1}{A^2 - A}$ *and let* (μ, b) *be a configuration that is not equivalent to* (μ^*, b^*)*. Then* (μ, b) *is not undominated neutrally stable.*

The proposition follows immediately from the following five lemmas.

First, Lemma 3 shows that dominated neutrally stable configuration must include more than one type in their support, and that the lowest type must be L_1 or L_2. Formally:

Lemma 3. *Let (μ, b) be a configuration such that b is a Bayesian-Nash equilibrium given μ. Let type $L_{k_1} \in C(\mu)$ be the smallest type in the population. Then:*

1. *Everyone defects (with probability 1) at any horizon weakly smaller than k_1.*
2. *Any type $L_k \neq L_{k_1}$ in the population defects (with probability 1) at horizon $k_1 + 1$.*
3. *If $k_1 < M$ and $\mu(L_{k_1}) = 1$ the configuration* is *not neutrally stable.*
4. *If $k_1 > 2$ and $p > \frac{1}{A-1}$ then the configuration is not dominated neutrally stable.*
5. *If $p > \frac{1}{A-1}$ then $L_{k_1} \in \{L_1, L_2\}$ and $\mu(L_{k_1}) < 1$.*

Proof.

1. It is common knowledge that all types are at least k_1. This implies that defecting when the horizon is at most k_1, defecting at all remaining stages is the unique strategy that survives iterations of eliminating dominated strategies, and thus all players must defect with probability 1 when the horizon is at most k_1 given any signal about the opponent.
2. Part (1) implies that defecting is strictly better than cooperating at horizon $k_1 + 1$ for agents of type higher than k_1.
3. Observe that if $k_1 < M$, then ϵ mutants of type L_{k_1+1} who play d_{k_1+1} and enter the population, would outperform the incumbents.
4. For a sufficiently small $\epsilon > 0$,ϵ mutants of type L_1 who enter the population (and play d_1) would achieve a higher payoff (for any $\delta > 0$) if:

$$(A-2) \cdot p > 1 - p \;\Leftrightarrow\; P > \frac{1}{A-1}$$

   This is because when type L_1 is identified, it is strictly dominating for his observing opponent to cooperate at all horizons strictly larger than 2. Thus, when being observed, L_1 mutants get at least $(A-2)$ fitness points more than L_{k_1} (as the opponent will cooperate for at least one more turn). When being unobserved, L_1 mutants obtain at most 1 point less than the L_{k_1}incumbents.

5. It is immediately implied by parts (3) and (4).

□

The following lemma shows that if everyone cooperates at all horizons strictly larger than $k_1 + 1$ in a dominated neutrally-stable configuration, then this configuration must be equivalent to (μ^*, b^*).

Lemma 4. *Let $\frac{A-1}{A} > p$. Let (μ, b) be an undominated neutrally stable strategy, and let type $L_{k_1} \in C(\mu)$ be the smallest type in the population such that $\mu(L_{k_1}) < 1$ and $k_1 \leq M - 2$. Denote the remaining types in $C(\mu)$ besides L_{k_1} as non-naive incumbents. Assume that all types in the population cooperate at all horizons strictly larger than $k_1 + 1$ when facing strangers. Then:*

1. *No one defects at horizon strictly larger then $k_1 + 2$ against any incumbent.*
2. *All non-naive incumbents play d_{k_1+1} against strangers or observed type L_{k_1}, and plays d_{k_1+2} against any non-naive observed incumbent.*
3. *No player in the population has type strictly larger than L_{k_1+2}.*
4. *The population only includes types $\{L_{k_1}, L_{k_1+2}\}$.*
5. $$\mu(L_{k_1}) = \frac{p(A-1) - 1 + \delta \cdot (c(L_{k_1+2}) - c(L_{k_1}))}{p(A-1)}$$
   (for any $p > \frac{1}{A-1}$, and no neutrally stable configuration exists if $p < \frac{1}{A-1}$).
6. *(μ, b) and (μ^*, b^*) are equivalent configurations.*

Proof.

1. We have to show that playing d_{k_1+2} is strictly better then an earlier defection against an observed non-naive incumbent. This is because defecting at horizon $k_1 + 3$ (defecting at horizon strictly larger than $k_1 + 3$) yields $A - 1$ (at least $2 \cdot (A - 1)$) less points than d_{k_1+2} against an unobserving opponent and at most 1 (2) more points than d_{k_1+2} against an observing opponent. Thus d_{k_1+2} is strictly better than defecting at horizon of at least $k_1 + 3$ if:
$$(1-p) \cdot (A-1) > p \;\Leftrightarrow\; (A-1) > A \cdot p \;\Leftrightarrow\; \frac{A-1}{A} > p.$$
2. By part (2) of the previous lemma all non-naive incumbents play d_{k_1+1} when facing strangers or observed L_{k_1}. It is immediate that d_{k_1+2} is strictly better than defecting at horizon of at most $k_1 + 1$ when facing an observed incumbent. By the previous part,

any incumbent with type strictly larger then L_{k_1+1} play d_{k_1+2} against observed non-naive incumbents. In order to complete the proof we have to show that all non-naive incumbents have type different then L_{k_1+1}. Assume to the contrary that: (I) all non-naive incumbents have type L_{k_1+1}; this implies that mutants of type L_{k_1+2}who play d_{k_1+2}against non-naive incumbents and d_{k_1+1} against strangers or naive incumbents, would outperform the incumbents; or (II) Some of the non-naive incumbents have type L_{k_1+1} while other incumbents have higher types; then for sufficiently small $\delta > 0$, the latter group outperform the former.

3. Assume to the contrary that there are players of type strictly higher than L_{k_1+2}. If there are also incumbents of type L_{k_1+1} then the previous part shows that both groups play the same on-equilibrium-path, and thus the agents with the strictly higher types must obtain strictly lower payoffs due to the cognitive costs. Otherwise, any best-reply mutant type L_{k_1+1} must play d_{k_1+1} against strangers and naive incumbents, and this implies that in any undominated configuration, non-naive incumbents cannot defect at horizons strictly higher than k_{1+2} when facing an observed mutant type L_{k_1+1}. This implies that such mutants would outperform the incumbents due to the cognitive costs.

4. This is immediate from the previous two parts.

5. In any balanced configuration the naive and the non-naive incumbents must have the same payoff. By repeating the calculation of Lemma 1, this can only hold if $\mu\left(L_{k_1}\right) = \frac{p(A-1)-1+\delta\cdot\left(c\left(L_{k_1+2}\right)-c\left(L_{k_1}\right)\right)}{p(A-1)} > 0$.

6. If $L_{k_1} = L_1$ then the previous parts imply that (μ, b) and (μ^*, b^*) are equivalent configurations. Assume to the contrary that $L_{k_1} = L_2$. We now show that ϵ mutants of type L_1 who invade the population would outperform the incumbents of type L_2 (and this immediately implies that the mutants also outperform the incumbents of type L_4, as the post-entree difference in the payoffs between the incumbents is $O\left(\epsilon\right)$). When facing an opponent of type L_2 or an unobserving opponent of type L_4, the mutants obtain one less point. When facing an observing opponent of type L_4, the mutants obtain $A-1$ more fitness points. Thus the mutants achieve a strictly higher payoff if:

$$\mu\left(L_2\right) + \mu\left(L_4\right)\cdot\left(1-p\right) < p\cdot\left(A-1\right)\cdot\mu\left(L_4\right),$$

$$\mu\left(L_2\right) + \mu\left(L_4\right) < p\cdot A\cdot\mu\left(L_4\right) \Leftrightarrow 1 < p\cdot A\cdot\mu\left(L_4\right) \Leftrightarrow \mu\left(L_4\right) > \frac{1}{p\cdot A}.$$

By the previous part:

$$\mu\left(L_4\right)=\frac{1-\delta\cdot\left(c\left(L_4\right)-c\left(L_2\right)\right)}{p\cdot\left(A-1\right)}>\frac{1}{p\cdot A}$$

for a sufficiently small δ.

□

Lemma 5. *Let $1-\frac{2\cdot A-1}{A^2-A}>p$. Let (μ,b) be an undominated neutrally stable strategy, and let type $L_{k_1}\in C\left(\mu\right)$ be the smallest type in the population such that $\mu\left(L_{k_1}\right)<1$ and $k_1\leq M-2$. Denote the remaining types in $C\left(\mu\right)$ besides L_{k_1} as non-naive incumbents. Let η be the mean probability that a random non-naive incumbent cooperates at all horizons strictly larger than k_1+1 when facing a stranger. Assume that $0<\eta<1$. Then:*

1. $\mu\left(L_{k_1}\right)\leq\frac{1}{A}$.
2. *If $p>\frac{1}{(A-1)^2}$, then $L_{k_1}=L_1$.*
3. $$\eta>\frac{(A-1)\cdot(1-p)-1}{(A-1)\cdot(1-p)}.$$
4. *η of the non-naive incumbents play d_{k_1+1} against strangers and the remaining $1-\eta$ play d_{k_1+2} against strangers.*
5. *When facing any incumbent, all types cooperate with probability 1 at all horizons strictly larger than k_1+3.*
6. *No player in the population has type strictly larger than L_{k_1+3}.*

Proof. □

1. The fact that there are incumbents who defect with with positive probability at horizons strictly larger than k_1+1 against strangers implies that early defection (at horizon strictly earlier than k_1+1) yields a weakly-better payoff than d_{k_1+1} against strangers. Early defection at horizon k_1+2 ($>k_1+2$) yields at least $A-1$ ($2\cdot(A-1)$) less fitness points against naive agents, and at most 1 (2) more points against non-naive opponents. This can hold only if:

$$\mu\left(L_{k_1}\right)\cdot\left(A-1\right)\leq\left(1-\mu\left(L_{k_1}\right)\right)\cdot 1$$

$$\mu\left(L_{k_1}\right)\leq\frac{1}{A}. \tag{2}$$

2. Assume to the contrary that $k_1 > 1$. Observe that ϵ mutants of type L_1 would outperform the incumbents of type L_{k_1}(and thus would outperform all the incumbents in the post-entry configuration) if:

$$p \cdot (A-1) \cdot (1 - \mu(L_2)) > \mu(L_2) \cdot 1$$

This is because the mutants of type L_1 earn at-least $A-1$ more points when their type is observed by a non-naive incumbent, they earn the same when their type is not observed by a non-naive incumbent, and they earn at most 1 less point when playing against a naive incumbent (type L_{k_1}). Thus the mutants would achieve a strictly higher payoff if:

$$p \cdot (A-1) > \mu(L_2) \cdot (1 + p \cdot (A-1)) \Leftrightarrow \frac{p \cdot (A-1)}{1 + p \cdot (A-1)} > \mu(L_2).$$

Substituting (2) yields:

$$\frac{p \cdot (A-1)}{1 + p \cdot (A-1)} > \frac{1}{A} \Leftrightarrow p \cdot A \cdot (A-1) > 1 + p \cdot (A-1) \Leftrightarrow p > \frac{1}{(A-1)^2}.$$

To simplify notation, we will assume $k_1 = 1$ in the following proofs (though they hold also for $k_1 \neq 1$ which may be possible for $p \leq \frac{1}{(A-1)^2}$).

3. Type L_1 gets $(L-1) \cdot A + 1$ points when playing against itself. A random player with a type different than L_1 who plays against L_1 gets at most $(L-1) \cdot A + 1 + 1$ when he observes his opponent's type, and an expected payoff of at most $\eta \cdot ((L-1)A + 2) + (1-\eta) \cdot ((L-2) \cdot A + 3)$. This implies that a necessary condition for other types to achieve a higher payoff (on average) when playing against L_1 than the payoff that L_1 gets against itself is (subtracting the equal amount of $(L-2) \cdot A + 1$ from each payoff):

$$A < p \cdot (A+1) + (1-p) \cdot (\eta \cdot (A+1) + 2 \cdot (1-\eta))$$

$$A < 1 + p \cdot A + (1-p) \cdot (\eta \cdot A + 1 - \eta) \Leftrightarrow A - \frac{1}{1-p} < \eta \cdot A + 1 - \eta$$

$$A - 1 - \frac{1}{1-p} < \eta \cdot (A-1) \Leftrightarrow 1 - \frac{1}{(A-1) \cdot (1-p)} < \eta$$

$$\frac{(A-1) \cdot (1-p) - 1}{(A-1) \cdot (1-p)} < \eta \tag{3}$$

If (3) does not hold, then the configuration cannot be naturally stable, because a

sufficiently small group of mutants with type L_1 who play d_1 would outperform the incumbents.

4. We show that when facing strangers, all types cooperate with probability 1 at all horizons strictly larger than 3. Assume to the contrary that there is a type who defects with positive probability against strangers at horizon $l > 3$. This implies that defecting at horizon l yields a weakly better payoff against strangers than d_3. This can occur only if:

$$\eta \cdot (1-p) \cdot (A-1) \leq (1-\eta) + \eta \cdot p.$$

This is because if $l = 4$ ($l > 4$), d_{k_1+2} yields $A-1$ (at least $2 \cdot (A-1)$) more points against non-observing opponents who cooperate at all horizons larger than 2, and it yields at most 1 (2) less points against any other opponents. This implies:

$$\eta \cdot (1-p) \cdot (A-1) \leq 1 - \eta \cdot (1-p) \Leftrightarrow \eta \cdot (1-p) \cdot A \leq 1\,(1-p) \Leftrightarrow \eta \leq \frac{1}{(1-p) \cdot A}.$$

Substituting (3) yields:

$$\frac{(A-1) \cdot (1-p) - 1}{(A-1) \cdot (1-p)} \leq \frac{1}{(1-p) \cdot A} \Leftrightarrow A \cdot ((A-1) \cdot (1-p) - 1) \leq A - 1$$

$$A \cdot (A-1) \cdot (1-p) - A \leq A - 1 \Leftrightarrow A \cdot (A-1) \cdot (1-p) \leq 2 \cdot A - 1$$

$$1 - p \leq \frac{2 \cdot A - 1}{A \cdot (A-1)} \Leftrightarrow p \geq 1 - \frac{2 \cdot A - 1}{A^2 - A}$$

and we get a contradiction to $p < 1 - \frac{2 \cdot A - 1}{A^2 - A}$. By part (2) of Lemma 3, all non-naive incumbents defect with probability 1 at any horizon of at most 2. This implies that η of the non-naive incumbents play d_2 against strangers and the remaining $1 - \eta$ play d_3 against strangers.

5. The proof repeats the same argument of part (1) of the previous lemma.

6. The proof repeats the same argument of part (3) of the previous lemma.

Lemma 6. *Let $1 > p > 0$. Let (μ, b) be a configuration, and let type $L_1 \in C(\mu)$ be the smallest type in the population ($\mu(L_{k_1}) < 1$). Denote the remaining types in $C(\mu)$ besides L_1 as non-naive incumbents. Let η be the mean probability that a random non-naive incumbent cooperates at all horizons strictly larger than $k_1 + 1$ when facing a stranger. Assume that $0 < \eta < 1$. Then (μ, b) cannot be undominated neutrally stable.*

Proof. Assume to the contrary that (μ, b) is undominated neutrally stable configuration.

1. *All players who play d_2 against strangers have type L_2.*
   Assume to the contrary that there is a type $L_{\tilde{k}}$ ($\tilde{k} > 2$) that plays d_2 with positive probability against strangers (and by the previous lemma it plays d_3 with the remaining probability). Consider the following configuration of mutants: (μ', b'): (1) $\mu' = \mu$, (2) for each $k \neq \tilde{k}$, $b'_k = b_k$, (3) for each $L_k \in \mathcal{L}$, $b'_{\tilde{k}}(k) = b_{\tilde{k}}(k)$, and (4) $b'_{\tilde{k}}(\emptyset) = d_3$. That is, the mutants have the same distribution of types as the incumbents, and they play the same except that mutants of type $L_{\tilde{k}}$ always play d_3 when facing strangers. Observe that such mutants would outperform the incumbents: mutants of type different than $L_{\tilde{k}}$ obtain the same payoff as their counter incumbents, mutants of type $L_{\tilde{k}}$ would achieve a strictly higher payoff when facing an unobserved opponent of type $L_{\tilde{k}}$ (pre-entry both d_2 and d_3 yielded the same payoff; post-entry there are a bit more early defectors and thus d_3 yield a strictly higher payoff), and would obtain the same payoff in all other cases. This implies that the configuration cannot be neutrally-stable.

2. $C(\mu) = \{L_1, L_2, L_4\}$. *Type L_1 always plays d_1, type L_2 always plays d_2, and type L_4 plays d_2 against observed L_1, d_3 against strangers and observed L_2, and plays d_4 against observed L_4.*
   By the previous lemma, there are no types strictly higher than L_4. By a similar argument to part (2) of Lemma 4, this implies that agents of type L_4 who play as the incumbents of type L_3 except that they play d_4 against an observed type L_3, would outperform agents of type L_3. Thus, type L_3 cannot be in the support of the population. The strategies that each type plays against other incumbents follow immediately from previous lemma and from the previous part of this lemma.

3. To simplify notation we characterize the frequency of each type in the case where the cognitive costs converge to 0 ($\delta \to o$). The arguments work very similarly (but the notation is more cumbersome) for for small enough δ. *Then:*

$$\mu(L_1) = \frac{1}{A + p \cdot (1-p) \cdot (A-1)^2}, \ \mu(L_2) = 1 - \frac{1 + A - p \cdot (A-1)}{A + p \cdot (1-p) \cdot (A-1)^2},$$

$$\mu(L_4) = \frac{1}{p \cdot (A-1) + 1}.$$

   Let $\mu_k = \mu(L_k)$. The fact that (μ, b) is a balanced configuration implies that types L_1 and L_2 should have the same payoff. Type L_2 obtains 1 more fitness point against types L_1 and L_2, the same payoff against an unobserving type L_4, and $A-1$ less points

against an observing type L_4. The balance between the payoffs implies:

$$(1-\mu_4) = \mu_4 \cdot p \cdot (A-1) \Leftrightarrow \mu_4 = \frac{1}{p \cdot (A-1)+1}. \tag{4}$$

Similarly, L_2 and L_4 should have the same payoff. Type L_2 obtain 1 less fitness point against opponent of type L_2, the same against observed type L_1, $A-1$ more points against unobserved type L_1, and the comparison against opponent of type L_4 depends on the observability: $A-2$ more points when both types are observed, 1 less point when both types are unobserved, 2 less points when only the opponent was observed, and $A-1$ more points when only the opponent was observing. Thus, the balance between the payoffs implies (taking into account also the cognitive costs):

$$(1-p) \cdot \mu_1 \cdot (A-1) + \mu_4 \cdot \left(p^2 \cdot (A-2) - (1-p)^2 + p \cdot (1-p) \cdot (A-1-2)\right) = \mu_2$$

$$(1-p) \cdot \mu_1 \cdot (A-1) + \mu_4 \cdot \left(p^2 \cdot (A-3) - 1 + 2p + \left(p - p^2\right) \cdot (A-3)\right) = \mu_2$$

$$(1-p) \cdot \mu_1 \cdot (A-1) + \mu_4 \cdot (p \cdot (A-3) - 1 + 2p) = \mu_2$$

$$(1-p) \cdot \mu_1 \cdot (A-1) + \mu_4 \cdot (p \cdot (A-1) - 1) = \mu_2 = 1 - \mu_1 - \mu_4$$

$$\mu_4 \cdot p \cdot (A-1) = 1 - \mu_1 \cdot (1 + (1-p) \cdot (A-1))$$

$$\mu_4 \cdot p \cdot (A-1) = 1 - \mu_1 \cdot (A - p \cdot (A-1))$$

$$\mu_1 \cdot (A - p \cdot (A-1)) = 1 - \mu_4 \cdot p \cdot (A-1)$$

$$\mu_1 = \frac{1 - \mu_4 \cdot p \cdot (A-1)}{A - p \cdot (A-1)}$$

Substituting (4) yields:

$$\mu_1 = \frac{1 - \frac{p \cdot (A-1)}{p \cdot (A-1)+1}}{A - p \cdot (A-1)} = \frac{\frac{1}{p \cdot (A-1)+1}}{A - p \cdot (A-1)}$$

$$\mu_1 = \frac{1}{(p \cdot (A-1)+1) \cdot (A - p \cdot (A-1))} = \frac{1}{A + p \cdot (1-p) \cdot (A-1)^2}.$$

This implies that:

$$\mu_2 = 1 - \mu_1 - \mu_4 = 1 - \frac{1}{(p \cdot (A-1)+1) \cdot (A - p \cdot (A-1))} - \frac{1}{p \cdot (A-1)+1}$$

$$\mu_2 = 1 - \frac{1 + A - p \cdot (A-1)}{(p \cdot (A-1) + 1) \cdot (A - p \cdot (A-1))} = 1 - \frac{1 + A - p \cdot (A-1)}{(p \cdot (A-1) + 1) \cdot (A - p \cdot (A-1))}$$

$$\mu_2 = 1 - \frac{1 + A - p \cdot (A-1)}{A + p \cdot (1-p) \cdot (A-1)^2}.$$

If any of the μ_i-s is not between 0 and 1 then no neutrally stable configuration exists.

4. *The configuration is not neutrally stable.*
   A direct algebraic calculation reveals that for sufficiently small $\epsilon, \epsilon' > 0$:

   (a) If $p < 0.5$ then ϵ "imitating" mutants with a configuration (μ', b') with $\mu'(L_1) = 1 - \mu(L_4) + \epsilon'$, $\mu'(L_2) = 0$, $\mu'(L_4) = \mu(L_4) - \epsilon'$, and $b' = b$ (play the same as the incumbents) outperform the incumbents in the post-entry population.

   (b) If $p > 0.5$ ϵ "imitating" mutants with a configuration (μ', b') with $\mu'(L_1) = 0$, $\mu'(L_2) = 1 - \mu(L_4) + \epsilon'$, $\mu'(L_4) = \mu(L_4) - \epsilon'$, and $b' = b$ outperform the incumbents post-entry population for sufficiently small ϵ.

□

A.3 Theorem 3 - Stable Configurations Near 0 and 1

Theorem. *3*

1. *Let* $0 \leq p < \frac{1}{(M-2)\cdot(A-1)}$. *Then there exists an undominated neutrally stable configuration* $\left(\tilde{\mu}, \tilde{b}\right)$ *where all players have type* L_M *and they play* d_M *against strangers and type* L_M, *and* d_{k+1} *against observed "mutant" type* $L_k < L_M$. *Moreover, any other neutrally stable configuration is equivalent to* $\left(\tilde{\mu}, \tilde{b}\right)$.

2. *Let* $1 \geq p > \frac{A-1}{A}$. *Then in any stable configuration there is a positive frequency of players of type* L_M, *and these players defect at all stages when observing an opponent of type* L_M.

Proof.

1. We begin by showing the stability of the configuration in which all players have type L_M and they defect at all stages. It is immediate that player best reply to each other. Consider ϵ mutants with type $k < L$ who invade the population. When facing incumbents, the mutants obtain 1 fitness point less when their type is unobserved, and

$(A-1)\cdot(M-k-1)-1$ more fitness points when their type is observed. Thus for sufficiently small ϵ and δ, the incumbents achieve a strictly lower payoff if:

$$(1-p) > p\cdot((A-1)\cdot(M-k-1)-1) \Leftrightarrow 1 > p\cdot(A-1)\cdot(M-k-1)$$

$$p < \frac{1}{(A-1)\cdot(M-k-1)}.$$

This implies that for any $p < \frac{1}{(M-2)\cdot(A-1)}$, the configuration is undominated neutrally stable.

Next we show that any non-equivalent configuration cannot be neutrally stable when $p < \frac{1}{A-1}$ (thus, if $\frac{1}{(M-2)\cdot(A-1)} < p < \frac{1}{A-1}$ then no neutrally stable configuration exists). If all players in the population have type L_M then they must all play d_M in any Bayesian-Nash equilibrium, as it is the unique serially strictly undominated strategy. Otherwise, let $L_k < L_M$ be the smallest type in the support of the population. If $k = M-1$, then it is immediate that agents with type L_M would outperform agents with type L_{M-1}. If $k \leq M-2$, then by repeating the arguments in Lemmas 3-6, one can see that no undominated neutrally stable configuration exists.

2. Let L_k be the highest type in the population. Let l be the largest horizon in which L_k begins defecting with positive probability against an observed opponent of the same type. If this probability is strictly less than 1, then by a similar argument to Part (1) of Lemma 6, the configuration is not neutrally stable (ϵ "imitating" mutants who differ only in that their L_k-s play d_l with probability 1 would achieve a strictly higher payoff in the post-entry population). Now, if $l < k$, then ϵ mutants of type L_k who play d_{l+1} (start defecting one stage earlier) against observed L_k, and play the same as the incumbents in all other cases, would outperform the incumbents of type L_k (and this implies they would outperform all incumbents):

$$p > (1-p)\cdot(A-1) \Leftrightarrow p\cdot A > (A-1) \Leftrightarrow p > \frac{A-1}{A}$$

(because the mutants obtain 1 more point when their observed L_k opponent observes their type, and they get at most $A-1$ less points when he does not observe their type; they obtain the same payoff against strangers and other observed opponents). From similar reasons, If $l = k < M$, then ϵ mutants of type L_{k+1} who play d_{k+1}against observed L_k, and play the same as the incumbents of type L_k in all other cases, would outperform incumbents of type L_k (and this implies they would outperform all incumbents) in any undominated neutrally stable configuration.

□

References

Alger, I., and J. Weibull (2012): "Homo Moralis-Preference evolution under incomplete information and assortative matching," Discussion paper, Toulouse School of Economics (TSE).

Andreoni, J., and J. Miller (1993): "Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence," *The Economic Journal*, 103(418), 570–585.

Bosch-Domenech, A., J. Montalvo, R. Nagel, and A. Satorra (2002): "One, two,(three), infinity,...: Newspaper and lab beauty-contest experiments," *The American Economic Review*, 92(5), 1687–1701.

Bruttel, L., W. Güth, and U. Kamecke (2012): "Finitely repeated prisoners dilemma experiments without a commonly known end," *International Journal of Game Theory*, pp. 1–25.

Camerer, C. (2003): *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C., T. Ho, and J. Chong (2004): "A cognitive hierarchy model of games," *The Quarterly Journal of Economics*, 119(3), 861–898.

Chong, J., C. Camerer, and T. Ho (2005): "Cognitive hierarchy: A limited thinking theory in games," *Experimental Business Research*, pp. 203–228.

Cooper, R., D. DeJong, R. Forsythe, and T. Ross (1996): "Cooperation without reputation: experimental evidence from prisoner's dilemma games," *Games and Economic Behavior*, 12, 187–218.

Costa-Gomes, M., and V. Crawford (2006): "Cognition and behavior in two-person guessing games: An experimental study," *The American economic review*, 96(5), 1737–1768.

Costa-Gomes, M., V. Crawford, and B. Broseta (2001): "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69(5), 1193–1235.

Crawford, V. (2003): "Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions," *The American Economic Review*, 93(1), 133–149.

Crawford, V. (2008): "Look-ups as the windows of the strategic soul: Studying cognition via information search in game experiments," *Perspectives on the Future of Economics: Positive and Normative Foundations, A. Caplin and A. Schotter, Eds. Oxford University Press*.

Crawford, V., and N. Iriberri (2007): "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica*, 75(6), 1721–1770.

Dekel, E., J. C. Ely, and O. Yilankaya (2007): "Evolution of Preferences," *Review of Economic Studies*, 74(3), 685–704.

Fershtman, C., K. L. Judd, and E. Kalai (1991): "Observable Contracts: Strategic Delegation and Cooperation," *International Economic Review*, 32(3), 551–559.

Frenkel, S., Y. Heller, and R. Teper (2012): "Endowment as a Blessing," .

Gill, D., and V. Prowse (2012): "Cognitive ability and learning to play equilibrium: A level-k analysis," .

Güth, W., and M. Yaari (1992): "Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach," in *Explaining Process and Change: Approaches to Evolutionary Economics*, ed. by U. Witt, pp. 23–34. The University of Michigan Press, Ann Arbor.

Haviland, W., H. Prins, and D. Walrath (2007): *Cultural anthropology: the human challenge*. Wadsworth Pub Co.

Heifetz, A., and A. Pauzner (2005): "Backward induction with players who doubt others' faultlessness," *Mathematical Social Sciences*, 50(3), 252–267.

Ho, T., C. Camerer, and K. Weigelt (1998): "Iterated dominance and iterated best response in experimental" p-beauty contests"," *The American Economic Review*, 88(4), 947–969.

Hyndman, K., A. Terracol, and J. Vaksmann (2012): "Beliefs and (In) Stability in Normal-Form Games," .

Jehiel, P. (2005): "Analogy-based expectation equilibrium," *Journal of Economic theory*, 123(2), 81–104.

Johnson, E., C. Camerer, S. Sen, and T. Rymon (2002): "Detecting failures of backward induction: Monitoring information search in sequential bargaining," *Journal of Economic Theory*, 104(1), 16–47.

Maynard-Smith, J. (1974): "The theory of games and the evolution of animal conflicts," *Journal of Theoretical Biology*, 47(1), 209 – 221.

Maynard Smith, J. (1982): "Evolution and the theory of games," .

McKelvey, R., and T. Palfrey (1995): "Quantal response equilibria for normal form games," *Games and Economic Behavior*, 10(1), 6–38.

Mohlin, E. (2012): "Evolution of theories of mind," *Games and Economic Behavior*, 75(1), 299 – 318.

Nagel, R. (1995): "Unraveling in guessing games: An experimental study," *The American Economic Review*, 85(5), 1313–1326.

Nagel, R., and F. Tang (1998): "Experimental results on the centipede game in normal form: an investigation on learning," *Journal of Mathematical Psychology*, 42(2), 356–384.

Neelin, J., H. Sonnenschein, and M. Spiegel (1988): "A further test of noncooperative bargaining theory: Comment," *The American Economic Review*, 78(4), 824–836.

Rapoport, A., and W. Amaldoss (2004): "Mixed strategies and iterative elimination of strongly dominated strategies: An experimental investigation of states of knowledge," *Journal of Economic Behavior & Organization*, 42(4), 483–521.

Robson, A. (2003): "The evolution of rationality and the Red Queen," *Journal of Economic Theory*, 111(1), 1–22.

Rosenthal, R. (1981): "Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox.," *Journal of Economic Theory*, 25(1), 92–100.

Selten, R., and R. Stoecker (1986): "End behavior in sequences of finite Prisoner's Dilemma supergames A learning theory approach," *Journal of Economic Behavior & Organization*, 7(1), 47–70.

Stahl, D. (1993): "Evolution of Smart-n Players," *Games and Economic Behavior*, 5(4), 604–617.

Stahl, D., and P. Wilson (1994): "Experimental evidence on players' models of other players," *Journal of economic behavior & organization*, 25(3), 309–327.

Stennek, J. (2000): "The survival value of assuming others to be rational," *International Journal of Game Theory*, 29(2), 147–163.

Swinkels, J. (1992): "Evolutionary stability with equilibrium entrants," *Journal of Economic Theory*, 57(2), 306–332.

Winter, E., I. Garcia-Jurado, and L. Mendez-Naya (2010): "Mental Equilibrium and Rational Emotions1," .